Structural Biology

SOLUTION SCATTERING FOR MODEL DISCRIMINATION – TESTS OF A METHOD

Chris Bailey-Kellogg*[1], Alan M. Friedman*[2], and Tim E. Williamson[2] (williate@purdue.edu)

[1]Purdue University Department of Computer Sciences, 1398 Computer Science Building, West Lafayette, IN, 47907
[2]Purdue University Department of Biological Sciences, 1392 Lilly Hall of Life Sciences, West Lafayette, IN, 47907

The onset of the post-genomic era has created a demand for the large-scale determination of protein structures. The traditional structural determination methods of x-ray crystallography and nuclear magnetic resonance (NMR), while producing high-resolution atomic models have experimental requirements that limit their application to large-scale studies; x-ray crystallography requires crystals of good diffraction quality, while NMR is limited to small soluble proteins. We are pursuing alternative methods employing x-ray solution scattering, a structural technique applicable to a broad range of conditions and sizes of macromolecules [Feigin & Svergun, 1987]. Solution scattering experiments involve passing an x-ray beam through a dilute solution of the target protein. Scattering intensity is then recorded as a function of the scattering angle. A spherically averaged diffraction pattern is produced from the random positions and orientation of particles in solution.

The creation of a structural model from solution scattering data requires several computational steps. After a curve of scattering intensity (I) vs. scattering angle (s) has been measured, a distance distribution curve displaying a probability distribution for all inter-atomic distances can be calculated using the indirect transformation method implemented in the program GNOM [Svergun, 1991], among other methods. This P(r) distribution can then be used to construct a dummy residue model of the scattered protein through *ab initio* restrained gas condensation, implemented in the program GASBOR [Svergun et al, 2001]. A transformation matrix can then be calculated that superimposes the dummy residue model onto a template structure so that a comparison of the reconstruction and a predicted model can be made based on a normalized spatial discrepancy test [Kozin & Sveregun, 2001]. Predicting the raw scattering curve of a protein based on a structural model is provided by the program CRYSOL [Svergun et al, 1995]. These tools provide three means of comparing experimental data to a predicted model of a protein structure arising from threading, *ab initio*, or other methods.

The ability to use solution scattering data to select the best models amongst a collection of predicted structures provides an experimentally based filter for determining which models are suitable for more intensive structure prediction computations and comparisons with other modes of experimental data. To meet this goal it is necessary to intelligently combine all of the above processes so that results from a large number of models can automatically be quantified. We are developing a system that allows for seamless integration of the aforementioned techniques and allows for model discrimination through the three separate metrics; scattering curve, probability density [P(r)] distribution, and normalized spatial discrepancy. The existence of three separate discrimination metrics allows for independent confirmation of the viability of each model.

We will demonstrate how our methods can be used to discriminate models in sets of CASP structure predictions and select the most correct models using the tri-metric approach, based on theoretical scattering curves predicted for the correct structure. We also will present data showing how these discrimination techniques can help construct and confirm a structural model based on experimental scattering data. Recently we have collected scattering data on the proteins SipA and Tfa at the APS BioCAT beamline, and have completed model reconstructions based on that data. These reconstructions were then tested against models created for each protein. We will describe our progress towards investigating whether solution scattering data can identify correct models through our tri-metric approach.

References:

Feigin, L. A., and D.I. Svergun, "Structure Analysis by Small-Angle X-Ray and Neutron Scattering" (1987) New York: Plenum Press. 335.

M. B. Kozin and D. I. Svergun, "Automated Matching of High and Low-Resolution Structural Models" Applied Crystallography (2001) 34, 33-41.

D. I. Svergun, "Determination of the Regularization Parameter in Indirect-Transformation Methods Using Perceptual Criteria" Applied Crystallography (1991) 25, 495-503.

D. I Svergun, et al., "CRYSOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates" Applied Crystallography (1995) 28, 768-773.

D. I. Svergun, et al., "Determination of Domain Structure of Proteins from X-Ray Solution Scattering" Biophysical Journal (2001) 80, 2946-2953.